Analysing Social Media as a Hybrid Tool to Detect and Interpret likely Radical Behavioural Traits for National Security

Pedro Cárdenas, Boguslaw Obara
Dept. Computer Science
Durham University
United Kingdom
pedro.cardenas-canto@durham.ac.uk
boguslaw.obara@durham.ac.uk

Georgios Theodoropoulos
Dept. of Computer Science and Engineering
Southern University of Science and Technology
China
georgios@sustec.edu.cn

Ibad Kureshi
Inlecom Systems, BVBA
Belgium
ibad.kureshi@inlecomsystems.com

***Abstract*—The study of National Security and its associated considerations is a sensitive and complex paradigm. It encapsulates both the protection of the territorial integrity and sovereignty of a state, as well as guaranteeing the security of its population. Known as Human Security, human-centred threats arising from radical activities need to be mitigated else they may escalate and have implications on National Security. The modern era has introduced further disruptive challenges, known as Hybrid Threats, that use non-traditional tools (Hybrid Tools) to intensify the impact of a likely threat. Social Media is a clear illustration of such tools, where the stability of the state and its people can be compromised by the dissemination of material. The ability to identify behaviour bordering on criminality within the deregulated world of Social Media is a Human Security imperative for governments. This paper follows on from our earlier work to detect affected National Security variables through the analysis of social media communication and trigger an alert when a likely threat is detected. As a result, a set of crisis interpretation processes are started to construe the event, such as radical behaviour analysis.**

This paper details the methodological approach to analyse one Hybrid Tool (Social Media) in order to identify likely instability scenarios based on the Human Security spectrum and therefore extract, detect and interpret dissimilar behavioural patterns that outline radical behavioural traits for National Security. The proposed methodology focuses on five steps, namely Instability Scenarios, Entity Extraction, Wordlists Creation, Content Analytics, and Data Interpretation.

***Index Terms*—National Security, Big Data and Radical Behaviour.**

I. Introduction

Societal instability issues across the globe tend to disrupt the fragile equilibrium of the state. Researchers and the government are able to monitor the state of a country by attempting to track societal markers such as Economic Security, Food Security, Health Security, Environmental Security, Personal Security, Communal Security and Political Security [1]. This in-turn contributes to the detection of larger problems that would affect National Security.

However, the detection and interpretation of security instabilities is non-trivial, as threats are evolving and depending on the source or domain include their own specific nuances. New unregulated domains, such as cyberspace, make the early detection and mitigation of threats a complex task for most governments [2]. Such threats are known as Hybrid Threats as they are not just confined to the digital realm but can spill out onto the streets. These newly shaped threats are aimed to affect societies at large, not just armies [3] and employ both conventional and unconventional methods such as military, economic, or technological, which can be used by different actors to disturb the human security components and destabilise the state. By creating confusion, inciting fear, blurring the institutional decision-making process, or undermine the confidence in the government, as described in [4], Hybrid threats can shake the foundations of government and society.

Different instruments/tools (also known as Hybrid Tools [5]) can be used to amplify the impact of such disruptive activities on National Security such as Propaganda, Domestic Media Outlets, Strategic Leaks, Political Parties, Paramilitary Organisations and Social Media among others [2].

The advent of the Internet has placed Social Media as a crucial asset to extract information from the virtual realm and strengthen a strategic analysis, since its core values lie firstly in the speed with which information travels, and secondly in the power that a set of clustered data offer unlike individual datum. This explains why the analysis of a set of tweets related to a specific topic unveil more information than a unique post on Twitter [6].

Social Media provides a platform that enables citizens to engage in discourse, and this plays a significant role in extracting the vox populi during a crisis event. Even during the lead up to such an event, the various text and contextual markers relating to National Security can be identified and mined to determine the extent to which the unfolding event will affect National Security [7].

Analysing and interpreting such information is a complex and challenging task, but it can provide useful insights that can be used by decision-makers to decide the best course of action to mitigate the disruptive situation. In line with this idea, in previous works [7]–[9] a holistic framework that utilises data

analytics to examine threats to National Security was proposed by considering three main components namely the Detonating Event, the Warning Period, and the Crisis Interpretation stages (see Figure 1).

The first stage (Detonating Event) is aimed at collecting those messages that embody the crowd reaction. The second stage (Warning Period) identifies which National Security variables were affected and consequently triggers an alert when the spotted incident trends to constitute a National Security threat.

Once the alert has been triggered the Crisis Interpretation stage initiates a set of computational techniques to analyse the digital content for societal characteristics linked to disruptive behaviours, such as coordination and cooperation, ideology, web insights [9] or radical behaviour. Principally the Crisis Interpretation process extracts insights from the discourse around the disruptive situation by considering the triggered alert as a milestone.

As part of the Crisis Interpretation stage, this paper presents a fine-grained methodology for the analysis of one Hybrid Tool (Social Media) to spot Hybrid Threats which are represented as the imbalance of the human security components, and therefore extract, detect and interpret a variety of radical behavioural traits for National Security, by considering five steps namely Instability Scenarios, Entity Extraction, Wordlists Creation, Content Analytics, and Data Interpretation.

The rest of the paper is organised into six sections. Section II delves deeper on hybrid tools. Section III outlines the analysis of radical behavioural traits. Section IV dissects the proposed system architecture. Section V illustrates the operationalisation of the methodology using two real incidents and Section VI concludes the paper delineating challenges and future work.

II. Hybrid Threats and Tools

Technology has placed the information environment into a dynamic arena where a set of heterogeneous actors tend to convey their ideologies and messages by using a combination of platforms to communicate an overall story or event. Within those actors, there might be citizens and non-state entities [12], who pursue their own interests and agendas, and in an unregulated virtual environment can lead to misinformation/disinformation [5]; resulting in the destabilisation of National Security.

Therefore, disinformation can be used as a strategy by the disruptive actors to mould people's behaviour using a mixture of real and digital world activities - hybrid threats [13]. The impact of these threats can be measured and applied to realworld scenarios by employing conventional and unconventional disruptive methods, spanning diplomatic, economic or technological contexts [4]. Hence, specific objectives/targets at a precise time [2] can be achieved, ranging from affecting critical infrastructures to creating confusion in order to strike the decision-making process of the state [4].

Hybrid threats are an integral part of modern conflicts/events leveraging disruptive tools to fuel such incidents. These tools or instruments of power also called hybrid tools [2], contribute to intensifying the impact of the incident and can adopt a wide range of forms such as Propaganda, Domestic Media Outlets, Social Media, Funding of Organisations or Strategic Leaks [2]. In this work, one Social Media instrument (the microblog Twitter) will be used to analyse different incidents aimed to detect and interpret radical behaviour that might affect National Security.

III. Detecting Hybrid Threats and Radical Behaviour

In all countries, managing national security is dealing with different challenges on a daily basis. Hybrid threats in itself is a complex challenges as they use multiple means tailored to take advantage of the vulnerabilities that society is facing [14]. The multi-faceted approach is what makes these kinds of threats difficult to detect [15].

According to [14] one way to detect a Hybrid Threat is by analysing the Political, Military, Economic, Social, Information and Infrastructure (PMESII) domains, which are closely related to the human security components proposed by [1] (Economic Security, Food Security, Health Security, Environmental Security, Personal Security, Communal Security and Political Security). Therefore one relevant component stakeholders might want to find answers to based on insights from data analysis is:

Q1. What sort of instability is the state dealing with based on the human security spectrum?

Social instabilities generate an environment where information can be used to undermine the confidence in the government, and therefore, the crisis can evolve until it affects the stability of the whole state. Cyberspace contributes to this complex scenario where individuals can create throw-away accounts to send their messages, taking advantage of apparent anonymity, the spread aggressive, violent and illegal viewpoints [16].

Calls of violence and violent views are a clear example of such behavioural patterns, where individuals try to influence social discourse through a variety of communication channels. Within the process of exchanging information via the Internet, the digital messages might nuance a set of different activities ranging from cyberbullying/victimisation, harassment, cyberstalking, gang violence [24], to radical expressions [18].

According to [18], a radical expression refers to an act linked to a violent reaction, and such demeanour can be understood by analysing two "behavioural markers" proposed by [19] namely Fixation and Leakage. The former concept (Fixation) refers to the trend to repeat constantly in a written message to a specific key term; whereas the second concept (Leakage) appertains to the intent to damage a particular target.

Both behavioural markers, Fixation and Leakage, need to spot the target that might be affected, which is why unveiling the associated entities/actors can provide more information about the incident. In this paper, three actors/entities are proposed as critical players because of their importance for National Security.

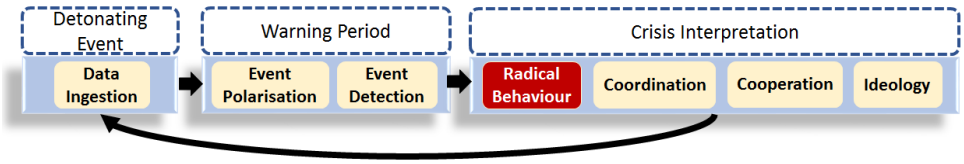

Fig. 1. Conceptual Framework for Social Movements Analytics for National Security

- Location; As described in [20] [21] settlements can be structured hierarchically according to their shape and population numbers, namely Conurbation, Metropolis, City, Large town, Small town, Large Village, Small Village, Hamlet and Isolated dwelling. A crisis event can emerge at several locations within the same country. Hence, a two-pronged strategy as described by [22] can be used to cluster such incidents, to wit: City-Level and Widespread events.
- People; For this paper, only those political leaders whose unexpected absence due to an attack or any other disruptive event and that might trigger a state of instability, such as President, Head of state, Prime Minister or Vice president, etc., will be considered.
- Strategic Facilities; Threats to National Security such as terrorism, tend to affect the national infrastructure and the balance of security and liberty of a state, as described in [23]. One way to deal with such disruptive events is by analysing those messages regarding strategic facilities such as airports, means of transport, schools, universities, roads or hospitals.

Hence, four questions pertinent to the above elements are:

Q2. Which entities are being mentioned during a disruptive incident?
Q3. Which entities can be affected due to their proximity to the affected entity?
Q4. What are the intentions that individuals express around the affected entities?
Q5. Has the incident disseminated at several locations?

Furthermore, the aforementioned behavioural markers consider violence as a critical element, as a disruptive incident has unfolded. However, the term violence can adopt multifarious definitions according to the context that is being analysed. Such consideration brings the chance to address violence from dissimilar perspectives that range from killing, doing harm [24], or rioting and looting [25], activities that identify a conflict. Therefore, it becomes necessary to analyse violence expressions within social media messages to disclose the different nuances of violence, which is why the classification proposed by [22] will be used, as it describes a set of phrases that can be found when analysing disruptive incidents.

In view of these performative acts that can be found during a crisis, two more questions that need to be addressed are:

Q6. Have people conveyed violent expressions during the crisis?
Q7. Which kind of violent expressions is being posted by individuals during the incident?

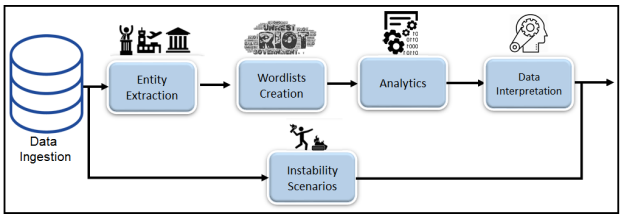

Fig. 2. Proposed System Architecture

Moreover, as described in our previous paper [7] *Coordination and Cooperation* is a stage part of the crisis interpretation process that refer to a cooperative system where individuals express actions during a disruptive incident aimed at reporting their needs that range from basic things such as water or shelter to objects that might comprise bombs, explosives or grenades. Hence, an additional question that needs to be addressed is:

Q8. Which type of necessities do people share on social media while a disruptive incident is taking place?

IV. System Architecture

The methodological approach is shown in Figure 2, and it comprises five stages (Instability Scenarios, Entity Extraction, Wordlist Creation, Content Analytics, and Data Interpretation) which are aimed at detecting and interpreting radical behavioural traits through the analysis of one hybrid tool (Social Media).

The foundations of the process are based on the analysis of different societal elements as described in Section III, which, in summary, enable to address eight questions that open up a multidimensional perspective since they provide key insights that can be used to interpret an incident.

A. Instability Scenarios (Q1)

A crisis event is an unstable situation that undermines the societal structure of a state and may lead to a disaster [26]. As described in Section III an instability tends to expose the vulnerabilities of the state. Therefore the analysis of the human security aspects over time can become a critical factor since a peaceful social movement can quickly escalate into a violent riot.

As a result, the projection of different scenarios that reveal which human security components have been compromised and that can be seen as a vulnerability by others will enable to understand the sort of instability the society is facing.

In line with this idea, in our previous work [8] a Deep Learning model was used to correlate the human security components, and the results suggested that individuals were involved towards a National Security issue when negative posts

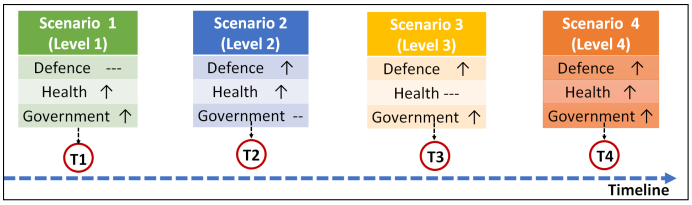

Fig. 3. Illustration of the proposed National Security scenarios. Scenario 1 illustrates that Health and Government have been affected in T1. Scenario 2 indicates that Defence and Health have been compromised in T2. Scenario 3 shows that Defence and Government have been put out of balance in T3. Scenario 4 depicts that Defence, Health and Government have been affected in T4.

that involved Defence, Health and Government have been disseminated over the Internet.

These three variables (Defence, Health and Government) can be used to create four different scenarios, as shown in Figure 3.

Each scenario illustrates which human security components are being affected, but the proposed levels do not follow a hierarchical structure since National Security aspects are weighted in a dissimilar way from country to country.

B. Entity Extraction (Q2 and Q3)

Entities are text expressions that contain special meaning such as names of people, locations or organisations [27]; however, data which flows within Social Media platforms present typographical mistakes since individuals tend to write their messages with a lack of capitalisation or by using short words to refer people's names, which in summary hinders the extraction of such terms.

Extracting such entities implies a complex process that requires as a first step to mine all major targets from the area where the disturbing incident is taking place (Knowledge Base Extraction). As described in Section III, there are three main actors/entities that play the core role from a National Security perspective, namely Locations, People and Strategic Facilities.

For this work the knowledge base Wikidata was used to extract all the forenamed actors related to the critical area. These included airports, universities, gas stations, power stations (Strategic Facilities), villages, human settlements, counties, towns, highways, streets (Locations) and heads of state (People), see Figure 4.

As a second step (Token Replacement) tweets are tokenised and classified according to the universal parts of speech code, as described in [28]. Then nouns are extracted because a typographic error can create confusion between an entity and a noun, and as described by [26] such grammatical errors downgrade the effectiveness of the conventional Named Entity Recognition techniques.

The third step (Semantic Matching) lies in performing a semantic match between the nouns extracted in the previous step (Token Replacement) against the entities extracted during the first process (Knowledge Base Extraction), in order to identify which entities are present in the analysed dataset (tweets). This process can be performed by considering a

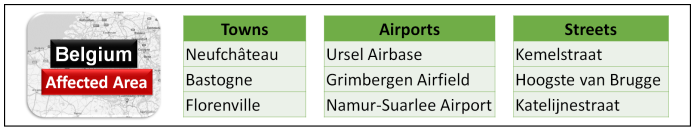

Fig. 4. Example of Location Extraction by querying Wikidata

string matching process where words are index by sound, in this case, a phonetic algorithm such as SoundEx was selected because it encodes words and characters by analysing their sound or pronunciation as described in [29], and in this way spelling errors can be detected.

However, one issue is that nouns are unigrams and entities (locations, people or facilities) are phrases that can act as single words (collocations) such as New York or Hong Kong. Hence, those entities that acted as collocations were joined as single words by using an underscore (new york –> `new_york`), then these set of new words were replaced in the original dataset (tweets), and afterwards, the Token Replacement and Semantic Matching processes were performed again. This step is crucial because, in such a way, nouns can represent either a word just as capitol or a complex phrase such as `hong_kong`.

The resulted nouns can be enriched or expanded. It was an enriched noun when collocations were corrected, whereas the noun was expanded when the word is an acronym, and it takes its original form, as shown in Table I.

TABLE I
EXAMPLE OF ENRICHED AND EXPANDED NOUNS

Nouns	Enriched Noun	Expanded Noun	Wikidata Description
new york	`new_york`	-	Global City
mong kok	`mong_kok`	-	Human Settlement
un	-	`united_nations_organisation`	International Organization
nato	-	`north_atlantic_treaty_organisation`	Military Alliance

According to [30] *Critical National Infrastructure* refers to those elements of infrastructure, which in case of being lost or compromised can impact on national security. In line with this idea, critical elements can be extracted by considering its proximity to the entities that have been identified from the previous step (Semantic Matching), such process can be performed by mining the information from a knowledge base such as Wikidata (see Figure 4).

Finally, all nouns that were recognised as entities have to be labelled by tags according to the knowledge base description (see Table I).

C. Wordlists Creation

As described in Section III, radical activity is linked to various violence nuances; however, detecting such behavioural patterns within social media messages represents a challenging task.

One way to deal with such a task is by creating a set of wordlists that contain nouns and verbs that enable the identification of specific actions and particular objects, such as weapons or means of transport. Therefore, this process can be dissected into three areas. The first area is focused on spotting

Communication Routes	People	Weapons	Supplies	Songs	Vehicles	Types of Waste
Road	Protestor	Teargas	Food	Anthem	Car	Rubbish
Street	Police	Bullet	Water	Song	Bike	Excrement
Thoroughfare	Soldier	Dynamite	Bread	Lyric	Scooter	Trash

TABLE II
OBJECT CLASSIFICATION DICTIONARY

those actions that reveal violent or non-violent actions; which is why the classification proposed by [22] can be used as a template to create the first dictionary (Dictionary of violence terms) which contains verbs and nouns related to both harsh and non-harsh activities.

The second area is aimed to create a dictionary (Dictionary of nouns) that classifies nouns/objects by its nature; for this work, seven different clusters were created, namely communication routes, people, songs, supplies, vehicles, weapons and types of waste. This dictionary is a core asset since it specifies whether the noun is a mean of transport, an explosive or a person (see Table II).

The last area is centred on creating a dictionary (Dictionary of verbs) which includes a list of verbs that describe a large group of activities; this is a key component because a range of intentions can be detected and in such a way policymakers can understand what sort of purposes people are conveying, such as occupy, assassinate, block or say, as shown in Table III.

D. Analytics (Q4 to Q8)

The interpretation of an incident needs the characterisation of reality which is a reflection of social behaviour, but measuring social perception requires the understanding of the set of actions performed by others, as described in [31]. Therefore actions can be described by verbs, which is why the interpretation process utilises verbs and nouns as the bedrock of this analysis.

In the first instance, this study considers the sentence breakdown where the basic parts are the subject, the verb and the object. Hence, the direct object is a noun phrase that expresses that an object/person is the recipient of an action verb.

Figure 5 depicts clear examples of noun phrases, where verbs express the intentions/actions, and the nouns represent the objects being acted upon. As a result, the direct object can be used to create a high-level summary of the analysed corpus. In a similar fashion to our last work [9], noun phrases are extracted, but in this case verbs are classified according to the dictionary of violence terms and the dictionary of verbs described above, and as far as nouns are concerned, they are categorised according to the entities and the dictionary of nouns previously mentioned.

In addition to the aforementioned method, a different technique can be used to enrich results such as word embeddings. This technique was selected because of its powerful way to represent words as vectors. This is important as word

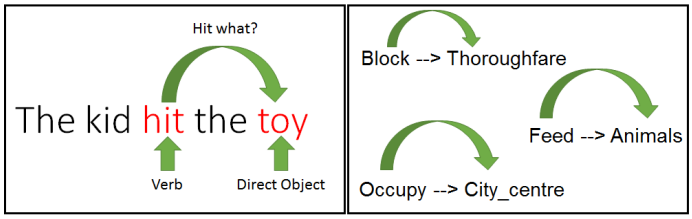

Fig. 5. Direct object sample

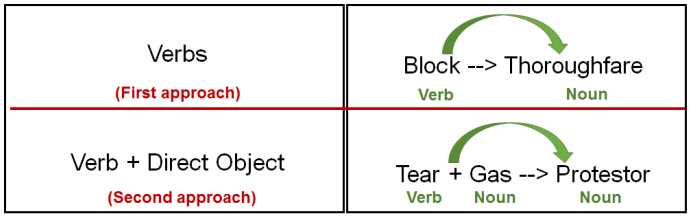

Fig. 6. Examining terms in the GloVe model to detect Radical Behavioural traits

vectors can show the semantic relationship between words [32]. However, when analysing radical events, scenarios are complex and varied since the necessities and targets that individuals pursue are different, which is why the semantic relationships between words will vary as well, and events need to be evaluated separately.

For this paper, the GloVe model was selected because it is an unsupervised learning algorithm focused on obtaining vector representations for words, as described in [33]. The GloVe model was trained by using the skip-gram method because the context in which words appear is a crucial factor, and in order to improve the quality of results a list of stopwords was removed from the datasets to create the vocabulary, and verbs were lemmatised.

Finally, examining the terms from the GloVe model was addressed in two ways. Firstly, verbs from the dictionary of verbs previously mentioned were used to find which word/noun reflect a semantic relationship, based on the specific context. Secondly, the verb and the direct object from the previous process were used to perform a mathematical operation (addition) between them, since those set of words can be considered as vectors, see Figure 6.

E. Data Interpretation

Interpreting data requires to organise the information in a specific manner that make it understandable. In line with this idea, the GloVe and direct object processes issued noun phrases that are formed essentially by two main elements,

Build Verbs	Verbs of Change of State	Verbs of Communication	Verbs of Contact by Impact	Fill Verbs	Want Verbs	Future Having	Verbs of Psychological State	Verbs of Social Interaction	Verbs of Creation	Verbs of Killing
Build	Crash	Explain	Bang	Block	Need	Feed	Affect	Argue	Build	Eliminate
Arrange	Break	Say	Beat	Bombard	Want	Give	Arouse	Combat	Assemble	Immolate
Churn	Shatter	Convey	Strike	Flood	Hope	Donate	Agitate	Clash	Bake	Liquidate

TABLE III
EXAMPLE OF VERBS THAT CAN BE USED TO INTERPRET ACTIONS. ADAPTED FROM [34]

Verb	Verb classification	Noun / Object	Object Classification	Violent / Non-violent	Behavioural Marker
Shoot	Verbs of Killing	Protestor	People	Violent	Leakage
Occupy	Verbs of Psychological State	Merida	Location	Non-violent	Fixation

TABLE IV
INTERPRETATION PROCESS EXAMPLE

namely a verb and a noun. As explained in Section III, verbs denote performing an activity/action but classifying such set of activities according to its meaning is a challenging task, which is why the categorisation proposed by [34] opens up a broad perspective.

Therefore, actions can be divided into eleven groups, as depicted in Table III, where activities such as kill, assassinate or block can disclose a manifest intention.

By contrast, nouns adopt different roles such as people, an object or a location; but by linking verbs that reveal a violent action with such nouns, new insights come to light and unveil radical behavioural traits and violent activity, see Table IV.

Regarding the coordination and cooperation traits, verbs that reflect the transfer of property will provide the evidence that people are looking for items or are offering them (e.g. bring pistol or need grenade). Such a list of verbs are based on the classification proposed by [34] and for this work verbs of creation, verbs of communication and future having will be considered to create the correspondent lexicon.

V. EXPERIMENTS AND VALIDATION

To demonstrate the validity and the robustness of the proposed methodological approach (see Figure 2), real disruptive incidents have be examined. The micro-bloging data from the protests in Hong Kong and the Ferguson riots in the United States of America have been used for validation purposes. These incidents were selected due to their diverse nature and the clear presence of radical behavioural.

For both incidents, Twitter's historical API was used to build the data corpus by extracting tweets based on hashtags regarded as trending ones. Then, data was cleansed and processed similarly to our previous works [8], [9], where retweets were selected as that is the mechanism of the microblog to spread feelings or ideas to reach new audiences [35].

The next essential step is analysing the sentiment fluctuations to identify whether negative feelings had the predominant role since they are the key ones during a disruptive incident, and afterwards, these group of messages were used to fuel the subsequent steps of the proposed methodological approach.

Figure 7 displays that in both cases (Hong Kong and Ferguson), negative feelings played the leading role, but in order to spot the point (date) when individuals felt attracted towards the disruptive incident and maybe heading to a situation where national security components can be compromised (tipping point), the Alert Mechanism proposed in our previous paper [8] was used, and it can be seen that in the Hong Kong case the tipping point was detected on September 27th, 2014, whereas the Ferguson riots case, the tipping point was detected on November 12th 2014.

A. Instability Scenarios (Q1)

When facing a crisis event, one question that must be answered is: What sort of instability is the state dealing with based on human security spectrum?

In order to tackle the former problem, tweets were classified into the ten aspects of human security components, as described in [1], to wit: Economy, Defence, Environment, Government, Health, Information, Life, Transport, People and Public Order.

This process was performed firstly by producing word embeddings to learn the context, and secondly, the embeddings were classified using a machine learning model. In this work, the word2vec algorithm was selected in view of its power to associate a vector with a word [32] [36], and a popular technique to categorise them - Gradient Boosting Machine-.

As time is a core factor in understanding the evolution of an incident, the percentage of each human security component was calculated per day, then to create the instability scenarios only three components were considered (Defence, Health and Government), as mentioned in Section IV.

*1) **The Hong Kong Protests**:* This event described a series of protests in 2014 when people began to manifest their discontent after a decision regarding the Hong Kongese electoral system was issued. The analysed period extends from September 26th —30th, 2014 as depicted in Figure 7.

For this case, Figure 8 illustrates that on the date the system triggered the tipping point (September 27th), the human security components that were affected were Defence and Government, whereas, during the next two days (September 28th-29th) when protestors took the streets the components changed to Health and Government, which suggests that people's health was compromised, and a day after (September 30th) the incident escalated since three components were affected namely Defence, Health and Government.

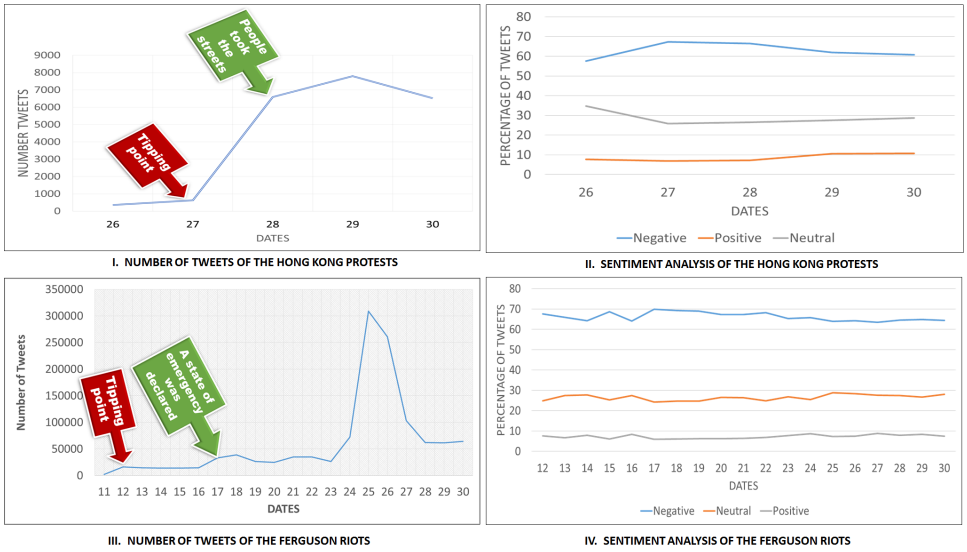

Fig. 7. Timeline of protests and Sentiment Analysis

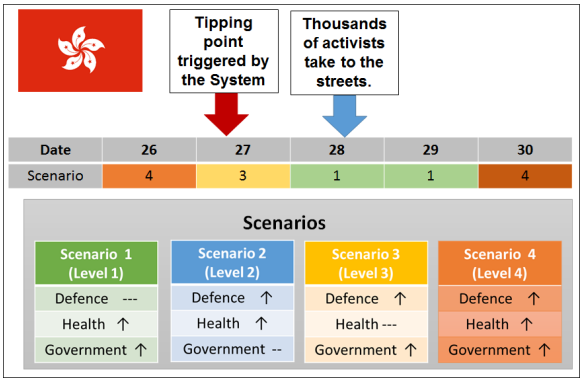

Fig. 8. Instability Scenarios in Hong Kong

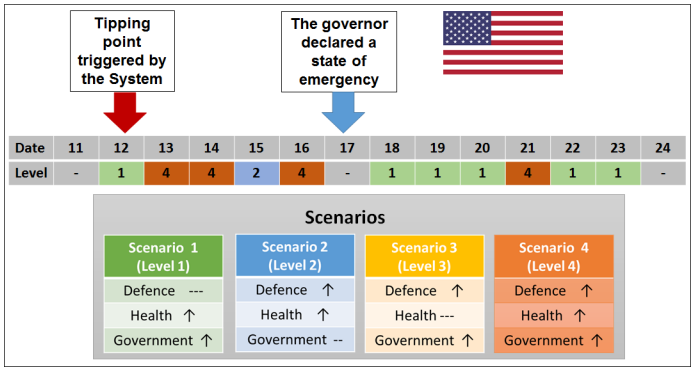

Fig. 9. Instability Scenarios in the USA (Ferguson riots)

*2) **The Ferguson Riots**:* These protests and riots began after the fatal shooting of a man by a police officer in 2014. The analysed period was from November 11th to 30th, as shown in Figure 7.

As displayed in Figure 9, the date where the tipping point was triggered (November 12th) two human security components were disturbed namely Health and Government, the next two days the scenario changed as three components showed disturbance (Defence, Health and Government), which suggests that societal problems escalated, and a day before the governor declared an estate of emergency (November 16th), the same scenario of three affected variables showed up.

B. Entity Extraction (Q2 and Q3)

After analysing instability scenarios over time, the next stage involves the extraction of main entities. This process represents a core stage because it enables the identification of people, locations and strategic facilities that can be considered as vital targets, as described in Section IV. Therefore, the entity extraction process was performed by considering a daily time frame, following the architecture depicted in Figure 2 described above.

*1) **The Hong Kong Protests**:*

Q2. This process extracted fifty-five entities, but for visualisation purposes, Figure 10 shows their distribution over time. These results suggest that on September 27th (Tipping point) individuals posted messages related to Cities, Human Settlements, Neighbourhoods and a Human being - President of the People's Republic of China-; this last entity might explain why the human security component -Government- is present in Scenarios 4 and 3 as shown in Figure 8.

A day after the tipping point was detected, individuals began sending messages about specific locations such as streets, and an activity ramp-up in entities such as Cities, Human Settlements and Neighbourhoods, which suggests that the incident was spread between different places.

Q3. Based on the entities extracted from the previous step, strategic targets were mined according to its proximity by considering the criteria proposed in Section IV. For exemplifying the results, entities that were close (10 km) to one of the extracted entities are shown in Figure 11.

*2) **The Ferguson Riots**:*

Q2. In this case, four entities were mentioned constantly

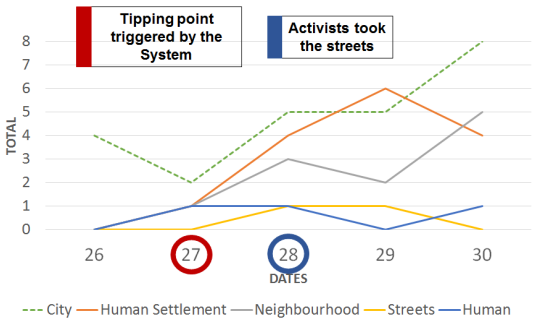

Fig. 10. Distribution of extracted entities (city, human settlement, neighbourhood, street and human) during the Hong Kong protests over time

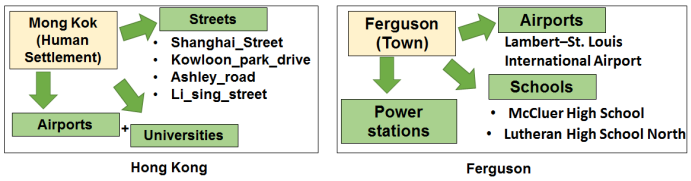

Fig. 11. Strategic Entities according to its proximity to affected areas

over time, namely Ferguson (town), Florissant (City), `Hereford_avenue` (road) and `Jay_nixon` (human being).

Q3. By considering Ferguson (town) as the axis of analysis, Figure 11 depicts some strategic facilities that can be extracted by considering a distance of 10 km from this location. It should be noted that these targets can provide a holistic view to deciding which entities need to be protected.

C. Analytics and Data Interpretation (Q4 to Q8)

As described in Section IV, a high-level summary is required to analyse, interpret and detect radical behavioural traits for National Security, so two different techniques were used to address that task, namely direct object and word embeddings.

1) *The Hong Kong Protests*:

Q4. Figure 12 depicts that the word embeddings process detected that individuals expressed their intention to occupy Central (neighbourhood) just a day before the tipping point was triggered by the system, by contrast, the direct object process identified that activists conveyed the intent to occupy Wanchai (neighbourhood), Mong Kok (human settlement) and Hong Kong (city) just the date when thousands of people took the streets.

Q5. Since individuals expressed their purpose to occupy multiple locations, this result suggests that the incident can be classified as a widespread event.

Q6. As described in Section IV, the Dictionary of violence terms contains verbs and nouns related to violent activities. As a result, the word embeddings process spotted turbulent activity when the tipping point was triggered because people described that they were suffering a tear gas attack and on subsequent days such violent activities continued as people conveyed gas attack, tear gas bullets and tear gas fire. On the other hand, the direct object process identified the violent acts as well, by the date when the streets were taken, as messages related to throwing canisters, clashing protestor, or a tear gas attack was published. It should be noted that messages regarding setting barricades, assembling barricades or occupying swaths were conveyed, although such activities are not considered as violent according to the classification proposed by [22], they can contribute to creating a plan for deciding the best course of action.

Q7. In this case, expressions linked to suffering a tear gas attack and throwing objects (canisters) are the violent expressions conveyed by individuals.

Q8. The Hong Kong case reflects that people sent tweets that denote the necessity of food, especially to donate noodle (September 30th).

2) *The Ferguson Riots*:

Q4. Figure 13 displays that the direct object process detected on November 12th (tipping point), individuals manifested their intentions to occupy Ferguson (town).

Q5. According to the analysed data corpus, individuals expressed the purpose to occupy one location, which is why it can be considered as a city-level event.

Q6. The Ferguson case shows that both processes, the direct object and word embeddings, disclose violent activity when the tipping point was triggered, as actions such as organising riot, organising boycott, looting, stealing or burning were posted (see the yellow-coloured line), and that kind of activity persisted during the following days, but also adding actions such as throwing excrement, throwing grenades or throwing dynamite (see Figure 13).

Q7. For this case, expressions linked to shooting protester or shooting thug was expressed before the tipping point, but a day after expressions such as shooting policeman or shooting black people were conveyed.

Q8. This particular event unveils a violent nature, but expressions such as offering dynamite, buying a weapon, buying ammunition, needing grenades, bringing grenades or building IED (improvised explosive device), suggests that social media worked as a mean to weaponise the incident.

VI. Conclusion and Future Work

This paper has introduced a novel technical methodology to analyse, detect and interpret radical behavioural traits by considering five steps, namely Instability Scenarios, Entity Extraction, Wordlists Creation, Analytics and Data Interpretation.

The proposed methodology enables a holistic analysis of the incident by:

- Creating scenarios to identify changes in human security components and,
- Extracting core entities relevant to the incident. These entities include:

(a) Locations

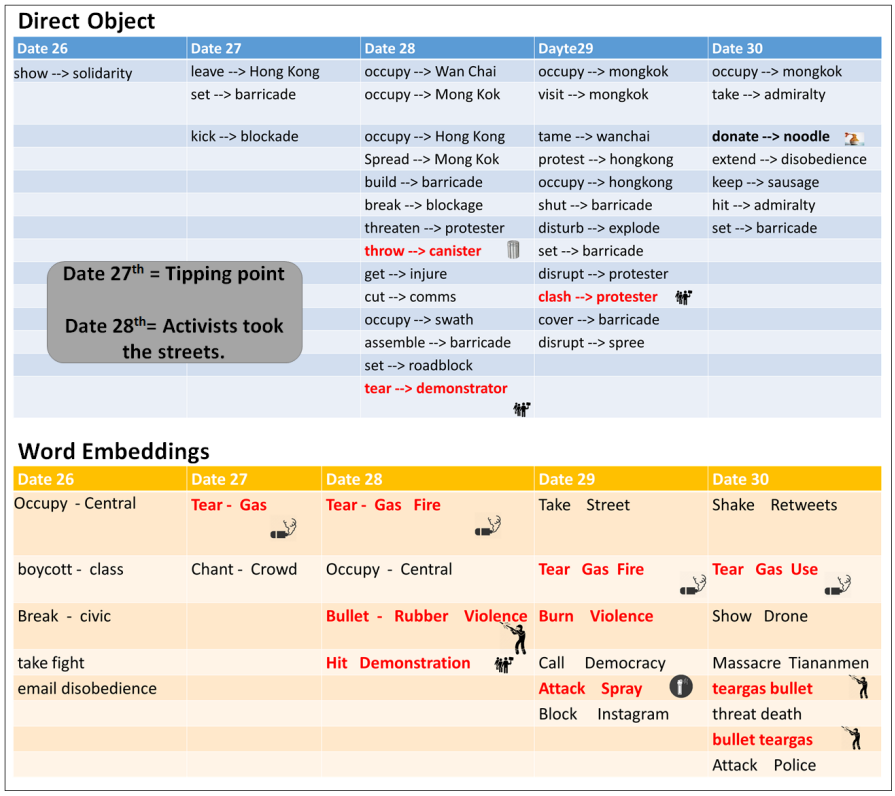

Fig. 12. Direct Object and Word Embeddings (Hong Kong)

Fig. 13. Direct Object and Word Embeddings (Ferguson riots)

(b) People
(c) Strategic facilities (determining dissemination at several locations) and,
(d) Those affected because of their proximity to the incident.

In addition, such analysis is complemented by detecting violent and non-violent expressions, and those instruments that can be used during the crisis, which can affect or complicate the way the event is evolving.

The validity and robustness of the proposed methodology have been demonstrated in the interpretation of radical behaviour during the protests in Hong Kong and the Ferguson riots in the USA. Results showed that violent expressions are different according to the nature of the incident and particular context.

Future work will analyse different disruptive cases for validation with the inclusion of different knowledge bases such as Wikipedia or DBpedia to improve the Entity Extraction process. An automated pipeline of each step of the process will also be devised for future real-time monitoring.

REFERENCES

[1] United Nations Development Program: Human Development Report. New York and Oxford: Oxford University Press, 22-33 (1994).

[2] Treverton G.F., Thvedt A., Chen A.R., Lee K. and McCue M.: Addressing Hybrid Threats, Swedish Defence University Center for Asymmetric Threat Studies, (2018).

[3] Treverton G.F.: The Intelligence Challenges of Hybrid Threats Focus on Cyber and Virtual Realm, Swedish Defence University Center for Asymmetric Threat Studies, (2018).

[4] Milo D., Draxler P., Klingova K., Misik M. and Pisko M.: Slovak Republic Hybrid Threats Vulnerability Study, Executive Summary, GLOBSEC, (2018).

[5] Svetoka S.: Social Media as a Tool of Hybrid Warfare, NATO Strategic Communication Center of Excellence, (2016).

[6] Heather J.W. and Llana B.: Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise, RAND Corporation, (2018).

[7] Cárdenas P., Theodoropoulos G., Obara B. and Kureshi I.: A Conceptual Framework for Social Movements Analytics for National Security. The International Conference on Computational Science, (2018).

[8] Cárdenas P., Theodoropoulos G., Obara B. and Kureshi I.: Defining an alert mechanism for detecting likely threats to National Security. IEEE International Conference on Big Data. USA (2018).

[9] Cárdenas P., Theodoropoulos G. and Obara B.: Web Insights for National Security: Analysing Participative Online Activity to Interpret Crises, IEEE International Conference on Cognitive Informatics and Cognitive Computing, Italy (2019).

[10] Tabansky L.: Cybered Influence Operations: Towards a Scientific Research Agenda, The Norwegian Atlantic Committee, (2017).

[11] Nissen T.E.: TheWeaponizationofSocialMedia: Characteristics of Contemporary Conflicts, Royal Danish Defence College, (2015).

[16] Correa D., Silva L.A., Mondal M., Benevenuto F., and Gummadi K. P.: The Many Shades of Anonymity: Characterizing Anonymous Social Media Content. In Proceedings of the 9th AAAI International Conference on Weblogs and Social Media (2015).

[12] Richterova J.: NATO and Hybrid Threats, Prague Student Summit, Model NATO, (2016).

[13] European External Action Service: EUDefence and Security Spring Series: Tackling new threats, (2019).

[14] Seeley B. and Shandra A.: Countering Hybrid Warfare: Conceptual Foundations and Implications for Defence Forces, Multinational Capability Development Campaign MCDC (2019).

[15] Monaghan S., Cullen P. and Wegge N.: MCDC Countering Hybrid Warfare Project: Countering Hybrid Warfare, Multinational Capability Development Campaign MCDC (2019).

[17] Meng Murat and Meng Seda: Violence and Social Media. Athens Journal of Mass Media and Communications- Volume 1 , Issue 3, 211-228 (2015).

[18] Meloy R., Hoffmann J., Guldimann A., James D.: The Role of Warning Behaviors in Threat Assessment: An Exploration and Suggested Typology . Behavioral Sciences and the Law 30, 256-279 (2012).

[19] Cohen K., Johansson F., Kaati L., Clausen M.J.: Detecting Linguistic Markers for Radical Violence in Social Media. Accepted Publ Terrorism Pol. Violence (2013).

[20] BBC Homepage, https://www.bbc.com/bitesize/guides/zyyvtyc/revision/4. Last accessed 9 Dec. 2018.

[21] Omblegya F. and Kernow T.: Settlements: Hierarchy and Settlement Categories. Cornwall Local Development Framework. (2011).

[22] Reed T., Belvin K., Hickman M., Kokolus M., Mendoza F., Phelps A. and Reisler K.: Open Source Indicators Program Handbook. The MITRE Corporation, USA (2017).

[23] Cabinet Office: The National Security Strategy of the United Kingdom, Security in an interdependent world. Available at: (https://assets.publishing.service.gov.uk). Accessed: 3 January 2019.

[24] Meng M., Meng S.: Violence and Social Media. Athens Journal of Mass Media and Communications No. 1 (3), pp. 211-227 (2015).

[25] Ray, L.: Shame and the City Looting, Emotions and Social Structure. The Sociological Review, 62(1), 117136, (2014).

[26] Castillo C.: Big Crisis Data: Social Media in Disasters and Time-Critical Situations.Cambridge University Press, (2016)

[27] Derczynski L., Maynard D, Rizzo G.,d , Van Erp M. , Gorrell G., Troncy R. , Petrak J. and Bontcheva K.: Analysis of named entity recognition and linking for tweets. Information Processing and Management, 51(2), 32-49 (2015).

[28] Taylor A.: A tidy data model for natural language processing using cleannlp. The R Journal 9(2):120 (2017).

[29] Bhatti, Z., Waqas, A., Ismaili, Imdad Ali, Hakro, Dil Nawaz and Soomro, W.J.: Phonetic based SoundEx and ShapeEx algorithm for Sindhi Spell Checker System, (2014).

[30] House of Lords, House of Commons, Joint Committee on National Security Strategy, National Security Strategy and Strategic Defence and Security Review, UK, (2016).

[31] Grezes, J., and de Gelder, B.: Social perception: understanding other peoples intentions and emotions through their actions, Social Cognition: Development, Neuroscience, and Autism, eds T. Striano and V. Reid. (Hoboken, NJ: Wiley-Blackwell), 6778, (2009).

[32] Chollet F. and Allaire J. J.: Deep Learning with R. Manning Publications Co., Greenwich, CT, USA, 1st edition, (2018).

[33] Pennington J., Socher R., and Manning C. D.: GloVe: Global vectors for word representation. In EMNLP, (2014).

[34] Levin B. English Verb Classes and Alternations: A Preliminary Investigation. University of Chicago Press, Chicago, (1993).

[35] Boyd D. and Golder S. and Lotan G.: Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. Hawaii International Conference on System Sciences, 1-10 (2010).

[36] Sugathadasa K., Ayesha B., de Silva N., Perera A.S., Jayawardana V., Lakmal D. and Perera M.: Synergistic Union of Word2Vec and Lexicon for Domain Specific Semantic Similarity. arXiv:1706.01967 (2017).